ARTICLE

The importance of mixed selectivity in complex cognitive tasks

Mattia Rigotti[1,2,3], Omri Barak[1]†, Melissa R. Warden[4,5], Xiao-Jing Wang[2,6], Nathaniel D. Daw[2,3], Earl K. Miller[4] & Stefano Fusi[1]

Single-neuron activity in the prefrontal cortex (PFC) is tuned to mixtures of multiple task-related aspects. Such mixed selectivity is highly heterogeneous, seemingly disordered and therefore difficult to interpret. We analysed the neural activity recorded in monkeys during an object sequence memory task to identify a role of mixed selectivity in subserving the cognitive functions ascribed to the PFC. We show that mixed selectivity neurons encode distributed information about all task-relevant aspects. Each aspect can be decoded from the population of neurons even when single-cell selectivity to that aspect is eliminated. Moreover, mixed selectivity offers a significant computational advantage over specialized responses in terms of the repertoire of input–output functions implementable by readout neurons. This advantage originates from the highly diverse nonlinear selectivity to mixtures of task-relevant variables, a signature of high-dimensional neural representations. Crucially, this dimensionality is predictive of animal behaviour as it collapses in error trials. Our findings recommend a shift of focus for future studies from neurons that have easily interpretable response tuning to the widely observed, but rarely analysed, mixed selectivity neurons.

Neurophysiology experiments in behaving animals are often analysed and modelled with a reverse engineering perspective, with the more or less explicit intention to identify highly specialized components with distinct functional roles in the behaviour under study. Physiologists often find the components they are looking for, contributing to the understanding of the functions and the underlying mechanisms of various brain areas, but they are also bewildered by numerous observations that are difficult to interpret. Many cells, especially in higher-order brain structures like the prefrontal cortex (PFC), often have complex and diverse response properties that are not organized anatomically, and that simultaneously reflect different parameters. These neurons are said to have mixed selectivity to multiple aspects of the task. For instance, in rule-based sensory-motor mapping tasks (such as the Wisconsin card sorting test), the response of a PFC cell may be correlated with parameters of the sensory stimuli, task rule, motor response or any combination of these[1,2]. The predominance of these mixed selectivity neurons seems to be a hallmark of PFC and other brain structures involved in cognition. Understanding such neural representations has been a major conceptual challenge in the field.

To characterize the statistics and understand the functional role of mixed selectivity, we analysed neural activity recorded in the PFC of monkeys during a sequence memory task[3]. Motivated by recent theoretical advances in understanding how machine learning principles play out in the functioning of neuronal circuits[4–10], we devised a new analysis of the recorded population activity. This analysis revealed that the observed mixed selectivity can be naturally understood as a signature of the information-encoding strategy of state-of-the-art classifiers like support vector machines[11]. Specifically we found that (1) the populations of recorded neurons encode distributed information that is not contained in classical selectivity to individual task aspects, (2) the recorded neural representations are high-dimensional, and (3) the dimensionality of the recorded neural representations predicts behavioural performance.

Dimensionality and mixed selectivity

The dimensionality of a neural representation in a given time bin is a property of a set of vectors, each of which represents the firing rates of N recorded neurons in one experimental condition. The pattern of activity encoded in each such vector can be thought of as a point in an N-dimensional space. Over a set of such points, the dimensionality we refer to is defined as the minimal number of coordinate axes that are needed to specify the position of all points (Supplementary Methods M.1). For example, if all points are on a line, then their dimensionality is one, as one appropriately aligned axis is sufficient to determine their position. The dimensionality of the neural representations recorded during an experiment is then the dimensionality generated by the patterns of activity observed in each of the different experimental conditions (for example, all combinations of sensory stimuli and behavioural responses).
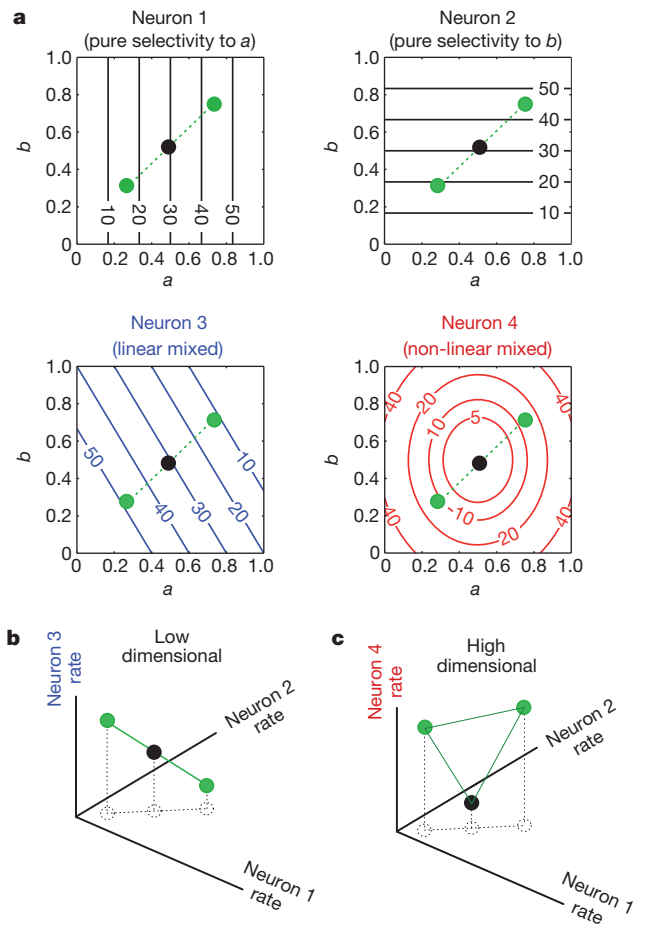
High-dimensional neural representations have the desirable property of allowing simple readouts such as linear classifiers to implement a large set of input–output relations. Model circuits that rely on such high-dimensional representations can generate very rich dynamics and solve complex tasks[5–10], and this same property is exploited in contemporary machine learning techniques such as support vector machines.

This dimensionality is related to the mixed selectivity of neuronal responses. This is because a set of neurons whose responses are selective only to individual task-relevant aspects, or even to linear sums of multiple aspects (linear mixed selectivity), can only generate low-dimensional representations. Higher dimensional representations can be produced by including neurons whose responses cannot be explained as a linear sum of aspect-related responses, that is, neurons with nonlinear mixed selectivity (Fig. 1a, b).

Mixed selectivity in a memory task

Monkeys were trained to remember the identity and order of presentation of two objects sequentially displayed on a screen. Their

---

[1]Center for Theoretical Neuroscience, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA. [2]Center for Neural Science, New York University, New York, New York 10003, USA. [3]Department of Psychology, New York University, New York, New York 10003, USA. [4]The Picower Institute for Learning and Memory & Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Bioengineering, Stanford University, Stanford 94305, California, USA. [6]Department of Neurobiology, Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut 06525, USA. †Present address: Department of Physiology, Technion Medical School, Haifa, 31096, Israel.

Figure 1 | Low and high-dimensional neural representations, and mixed selectivity. **a**, Contour plots of the responses (spikes per s) of four hypothetical neurons to two continuous parameters that characterize two task-relevant aspects (*a*,*b*, varying between 0 and 1) corresponding to relevant stimulus features (for example, contrast and orientation). Neurons 1,2 are pure selectivity neurons, selective to individual parameters (*a* and *b*, respectively). Neuron 3 is a linear mixed selectivity neuron: its response is a linear combination of the responses to parameters *a* and *b*. Neuron 4 is a nonlinear mixed selectivity neuron: its response cannot be explained by a linear superposition of responses to the individual parameters. The green circles indicate the responses to three sensory stimuli parameterized by three *a*,*b* combinations. **b**, The responses of the pure and linear mixed selectivity neurons from **a** in the space of activity patterns (the axes indicate the firing rates of the neurons) elicited by the three stimuli indicated by the green circles in **a** lie on a line, therefore spanning a low-dimensional space. **c**, As in **b**, with the third neuron being the nonlinear mixed selectivity Neuron 4 in **a**. The representations of the stimuli lie on a plane, no longer being confined on a line. This higher dimensionality has an important role when the activity is read out by linear classifiers, because they can only separate the input space into classes that are separable by a plane (in general by a hyper-plane). This limits the implementable classifications (See Supplementary Section S.1). For example, in **b** it is impossible for any linear classifier to respond to the darker central circle and not to the other two. But it is possible in **c**, for instance for a linear classifier corresponding to an appropriately positioned horizontal plane.

memory was then tested after a delay (the two-object delay period) either through a recognition or through a recall task, which were interleaved in blocks of 100–150 trials (Fig. 2 and refs 3, 12 for more details).

We analysed the activity of 237 lateral PFC neurons (area 46) recorded in two monkeys during the trial epochs that precede and include the two-object delay period. In these epochs, each trial was characterized by a condition defined by three relevant aspects: the identity of the two visual objects and the task type. The first cue was randomly selected from a predefined set of four objects. The second cue was randomly chosen among the three remaining objects.
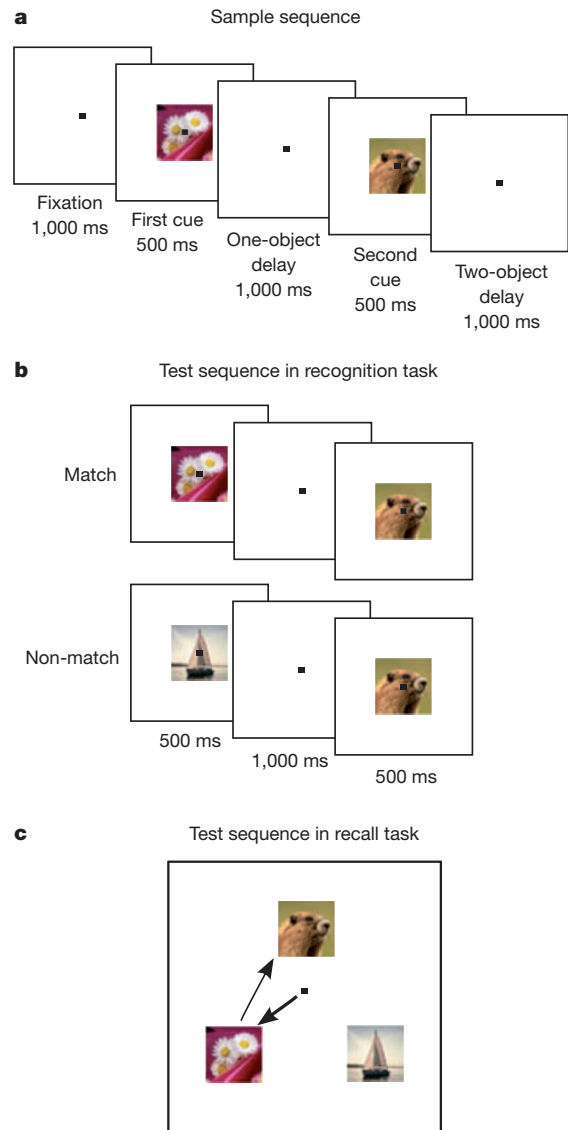
Figure 2 | Behavioural task from ref. 3. **a**, Sample sequence: each trial began when the monkeys grasped a bar and achieved central fixation. A first sample object was followed by a brief delay (the one-object delay), then a second sample object (different from the first sample object), then another delay (the two-object delay). **b**, Recognition task: the sample sequence was followed by a test sequence, which was either a match to the sample sequence, in which case the monkeys were required to release the bar, or a non-match, in which case the monkeys were required to hold the bar until a matching sequence appeared. **c**, Recall task: the sample sequence was followed by an array of three objects that included the two sample objects. Monkeys were required to make a sequence of saccades in the correct order to the two sample objects. Recognition and recall task trials were interleaved in blocks of 100–150 trials.

Finally, the task type was either recognition or recall. More generally, for other experimental protocols, every situation (condition) would be characterized by the values of a set of discrete or continuous variables. We refer to them as to the task-relevant aspects.

The analysis of recorded single-neuron responses shows that the majority of neurons are selective to at least one of the three task-relevant aspects in one or more epochs[3]. A large proportion of neurons, moreover, show nonlinear mixed selectivity (Supplementary Section S.2). Figure 3a–d show two examples of nonlinear mixed selectivity neurons. Figure 3a shows a cell that is selective to a mixture of cue 1 identity and task type: it responds to object C when presented as a first cue, more strongly during the recognition task. The neuron of Fig. 3c, d is mostly selective to objects A and D when presented as
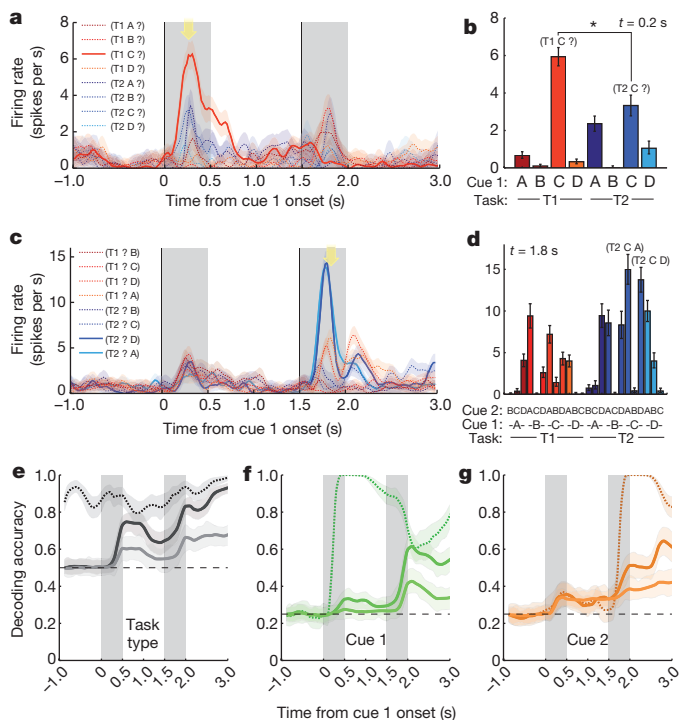
Figure 3 | Mixed selectivity in recorded single-cell activity and population decoding. a, Average firing rate of a sample neuron (Gaussian smoothing with 100-ms window, shaded area indicates s.e.m.). Colours denote different combinations of task type and sample cues (condition), indicated in parenthesis (task type, first cue, second cue). The '?' indicates that cue 2 identities were averaged over. This neuron preferentially responds to object C as first cue in task 1 blocks (recognition task). **b**, Peri-condition histogram (PCH): average firing rate in a 100-ms time bin ($\pm$ s.e.m.) at the yellow arrow in **a** for different conditions. The response to object C as first cue is significantly different for the two task types ($P < 0.05$, two-sample t-test). **c, d**, Same as **a, b**, for a different neuron with preference for object A and D as second objects during task 2 trials (recall task). **e–g**, Comparison of population decoding accuracy for task type (**e**), cue 1 (**f**) and cue 2 (**g**) before (dashed) and after (solid) removing classical selectivity. Dashed lines: average trial-by-trial cross-validated decoding accuracy of the decoder reading out the firing rate of 237 neurons in different independent time bins. Curves represent the average decoding accuracy over 1,000 partitions of the data into training and test set (shaded areas show 95% confidence intervals). Horizontal dashed lines indicate chance level. Solid lines: decoding accuracy after the removal of classical selectivity for 237 (bright) and 1,000 resampled neurons (dark) (see Supplementary Methods M.6). **e**, Accuracy in decoding task type from neurons whose selectivity to task type was removed. The decoding accuracy is initially at chance level, but steadily grows above chance level as the complexity of the task and the number of conditions increases. **f, g**, Analogous plots for the decoding accuracy of cue 1,2 identity, when instead selectivity to cue 1,2 was removed.

second stimuli, but only during the recall task and when they are preceded by object C.

Information encoded by mixed selectivity

A neuron is conventionally said to be selective to a task-relevant aspect if it responds differentially to the values of the parameters characterizing that aspect. A neuron is, for instance, selective to task type, if its average responses in recall and recognition task trials are significantly different. The operation of averaging over all conditions corresponding to a particular task type (for each task type there are 12 possible combinations of the sample visual objects) may, however, result in discarding important information. The responses in individual conditions could encode information about task type through the nonlinear interactions between the cue and the task type aspects, which manifest themselves as nonlinear mixed selectivity. This suggests that if nonlinear mixed selectivity is sufficiently diverse across

neurons, the information about task type could be extracted from the covariance between neuronal responses across different conditions, even when individual neurons are not 'classically' selective to task type (that is, the average responses to recall and recognition tasks are not significantly different). Information could in other words be distributed across the neural population, even when it is not present in individual cells (see also refs 13, 14 for recent discussions on distributed codes.)

To show that this is indeed the case, we manipulated the data to remove the classical selectivity to a given task aspect and then tested whether that task aspect could still be decoded from the population. Classical selectivity to a given task-relevant aspect is removed from every recorded neuron by adding noise that equalizes average responses, preserving the differences between the individual conditions (Supplementary Methods M.3). In Fig. 3e–g we compared the cross-validated trial-by-trial population-decoding accuracy before and after removing classical selectivity. Neurons that were not recorded simultaneously were combined as pseudo-simultaneous population activity patterns as explained in Supplementary Methods M.5. The temporal correlations that were neglected with this procedure do not seem to appreciably affect the decoding accuracy (Supplementary Section S.4).

Before removing classical selectivity, the maximal cross-validated decoding accuracy peaks close to 100% for all task-relevant aspects (Fig. 3e). Both the identity and the temporal order of the visual objects could also be decoded (see Supplementary Section S.6). Crucially, all task-relevant aspects can be decoded even when classical selectivity is removed. The first panel of Fig. 3e shows the accuracy of decoding task type from the intact population and after removing classical selectivity to task type from all neurons. Note that removing classical selectivity causes a larger drop of decoding accuracy in the early epochs of the trial. As the trial progresses and more visual cues are memorized (that is, the task becomes more complex), the accuracy progressively increases towards the values of the intact population. Moreover, the decoding accuracy increases as the number of neurons read out by the decoder increases. We estimated the decoding accuracy for larger neural populations by resampling the recorded neurons and randomly relabelling the identities of the visual objects, so as to obtain responses whose activity have the same statistics as the recorded ones. For example, a new neuron could be obtained by assigning the activity of a recorded neuron in response to objects A, B, C, D to the trials in which the objects were B, D, A, C (see Supplementary Methods M.3 and M.6 for more details). Similar results hold after removing the classical selectivity to cue 1 and cue 2 (Fig. 3f, g), or when we removed the classical selectivity by subtracting from the neural activity the linear mixed selectivity component (Supplementary Section S.3).
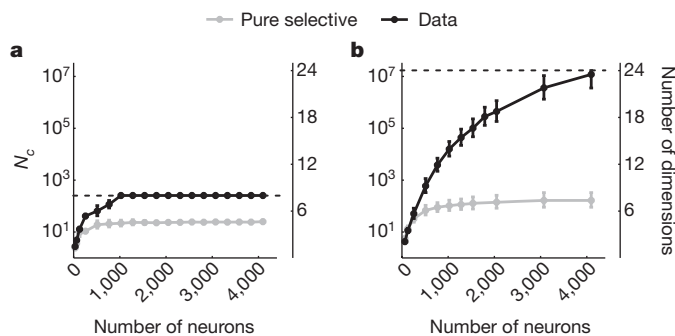
Neural representations are high–dimensional

To verify that the observed nonlinear mixed selectivity and the diversity of the neural responses are a signature of the high-dimensionality of the neural representations, we set out to quantify the dimensionality of the recorded activity. As this is notoriously difficult in the presence of noise (see Supplementary Section S.10 and ref. 15), we adopted a novel strategy that exploits the relation between dimensionality and the performance of a linear classifier reading out the neural activity (Supplementary Section S.7). Our method relies on the observation that the number of binary classifications that can be implemented by a linear classifier grows exponentially with the number of dimensions of the neural representations of the patterns of activities to be classified (Supplementary Methods M.1). Hence dimensionality can be estimated by counting the number of binary classifications that can be implemented by a linear classifier. The exponential dependence on the dimensionality implies that the number of implementable classifications can vary over several orders of magnitude, allowing for a huge computational advantage of high-dimensional over low-dimensional representations.

In general, a binary linear classifier reads out the activity of the recorded neurons and generates an output that can be either 0 or 1 depending on whether the sum of its weighted inputs is below or above a threshold. The set of weights determines the specific classification that is implemented. In our analysis, the classifier is trained to consistently generate the same output for all recorded trials corresponding to the same experimental condition. The output can be different for the c different conditions (here, c equals 24 possible combinations of values for the three task aspects), for a total of 2^c possible binary classifications or, equivalently, mappings from the c conditions to the binary outputs. Among all possible binary classifications we estimated through cross-validation how many are implementable by a linear classifier of the recorded patterns of activity (Supplementary Methods M.7). The total number of implementable binary classifications N_c is related to the estimated dimensionality d of the representations through the expression $d = \log_2 N_c$, when the number of inputs is sufficiently large (Supplementary Methods M.1 and Supplementary Section S.7). Accordingly, the dimensionality is bounded by the total number of distinct conditions c.

Figure 4 shows both N_c and d as a function of the number of neurons N read out by the classifier for two different neural representations. The first neural representation is given by the recorded PFC neurons. For values of N larger than the number of recorded cells, we computed the performance by introducing additional resampled neurons as previously described for the decoding analysis (Supplementary Methods M.6), after verifying that this does not introduce additional artificial dimensions to the neural representations (Supplementary Section S.8). The second neural representation is generated from simulated pure selectivity neurons that encode only one aspect of the task at a time (Supplementary Methods M.4). We computed N_c for these neural representations during two task epochs: the one-object (Fig. 4a) and the two-object delay (Fig. 4b). For both epochs, N_c grows with the size of the neural population N and it

saturates near the value that corresponds to the maximal dimensionality. The asymptotic value of N_c is always larger for the recorded representations than for the pure selectivity representations. The difference is several orders of magnitude for the two-object delay.

The ability to implement such a large number of classifications is due to the diversity of nonlinear mixed selectivity responses, which often results in seemingly disordered representations and response properties that are not easily interpretable. However, it is important to note that high-dimensional representations could also originate from more 'orderly' responses, in which each neuron behaves as a 'grandmother cell' that responds only to a single experimental condition—in our case to one out of 24. We ruled out this scenario in our data by verifying that PFC representations are rather dense (Supplementary Section S.11) and that the sparsest responses are not a major contributor to the observed high dimensionality (Supplementary Section S.19).

Dimensionality predicts animal performance

High-dimensional neural representations encode information that is nonlinearly mixed in a format that is suitable for local neural processing. Therefore high dimensionality could be important for the dynamics of neural circuits that eventually generate behaviour[4–10] (see also Supplementary Section S.1 for an illustration of the computational
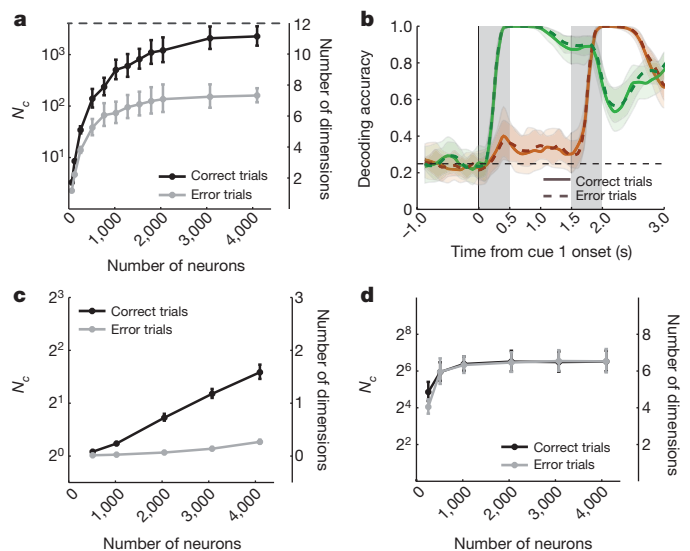
Figure 5 | The dimensionality of the neural representations predicts animal behaviour. a, Same plot as Fig. 4b, with the difference that the analysis is restricted to the recall task and the two curves represent the number of implementable binary classifications of the recorded activity in the case of correct (black) and error (grey) trials. For the correct trials the number of implementable classifications corresponds to a dimensionality that is close to maximal ($d = 12$, dashed line). In the error trials the dimensionality drops significantly. **b**, The identity of the two cues can still be decoded in the error trials: decoding accuracy as in Fig. 3 in the correct (continuous lines) and error trials (dashed lines) for the identity of cue 1 (green lines) and cue 2 (orange line). The correct cue identities are perfectly decoded also during error trials. The error bars (shaded areas) are computed as in Fig. 3e–g. **c**, **d**, Contribution of nonlinear and linear mixed selectivity to the collapse in dimensionality observed in the error trials. **c**, After removing the linear component of mixed selectivity from the response of each neuron, the dimensionality is estimated as in **a**. The dimensionality in the correct trials (black line) is still significantly higher than in the error trials (grey line). **d**, Same as in **c**, but after the nonlinear component of mixed selectivity is subtracted from each neuron. The two curves are not significantly different, indicating that the nonlinear component of mixed selectivity is responsible for the collapse in dimensionality. These analyses were carried out on a subset data set of 121 neurons that were recorded in as many correct trials as error trials during the recall task. **a**, **c** and **d**, Error bars are 95% confidence bounds estimated as detailed in Supplementary Methods M.7.

Figure 4 | The recorded neural representations are high-dimensional. Number of implementable binary classifications, N_c, (left ordinate, on logarithmic scale) and dimensionality of the inputs (right ordinate, linear scale) for varying number of neurons of the population read out by a linear classifier. **a**, The black trace represents the number N_c of implementable binary classifications of the vectors of recorded mean firing rates in the 800-ms bin in the middle of the one-object delay period. In this epoch a trial is defined by one of $c = 8$ different conditions, corresponding to all the combinations of task type and cue 1 objects. N_c reaches the value that corresponds to the maximal dimensionality $d = 8$ (indicated by the dashed line). The grey line shows N_c when the neural representations contain only the responses of artificially generated pure selectivity neurons with a noise level matching that of the data (See Supplementary Methods M.4). **b**, Same plot computed over the 800-ms bin in the middle of the two-object delay period. The advantage of the recorded representations over the pure selectivity neurons is huge. For the recorded data (black line) N_c reaches 2^{24}, the value that corresponds to the maximal dimensionality $d = 24$, given by all possible combinations of cue 1 object, cue 2 object and task type are 24 (dashed line). On the other hand, representations based on pure selectivity (grey line) generate less than 8 dimensions. Error bars are 95% confidence bounds estimated as detailed in Supplementary Methods M.7. See Supplementary Section S.20 for this analysis during the test epochs.

advantages of high-dimensional over low-dimensional representations). If this is the case, we should observe a correlation between the dimensionality of the neural representations and the performance of the animal. In particular, the dimensionality should decrease in error trials.

We tested this prediction by analysing the errors made by the monkeys in the recall task (there were too few errors in the recognition task). Figure 5a shows the number of implementable binary classifications and the dimensionality as a function of the number of neurons in the input in the 800-ms time bin in the middle of the two-object delay period, as in Fig. 4b, but only during the recall task and separately for correct and error trials. As predicted, the dimensionality decreases in the error trials. We ruled out that the decrease could be explained by a difference in the number of trials (for the 121 neurons with enough error trials, the difference between the number of correct and error trials is less than 4%), a change in average firing rate (Supplementary Section S.13), in the variability of single neuron activity (Supplementary Section S.14) or in the coding level of the activity (Supplementary Section S.15).

Remarkably, the identity of the two visual cues could still be decoded with high accuracy in the error trials. Figure 5d shows that the cross-validation performances of the decoder of Fig. 3 trained on correct trials and tested on a hold-out set of correct (continuous line) and error trials (dashed lines) are indistinguishable, demonstrating that when the monkeys make a mistake, cue identities are encoded with the same strength as in the correct trials. We verified that this correspondence is not a ceiling effect due to the population decoder saturating at high performance (Supplementary Section S.16).

These results indicate that the collapse in dimensionality observed in the error trials is not due to a failure in coding or remembering the sensory stimuli. Our hypothesis is that nonlinear mixed selectivity, which underlies high dimensionality, is important for the generation of the correct behavioural response. In the error trials this component of the selectivity is disrupted, leading to a collapse in dimensionality, which impairs the ability of downstream readout neurons to produce the correct response.

To test this hypothesis, we quantified the contribution to the collapse in dimensionality of the nonlinear mixed selectivity component of the neuronal response. The nonlinear mixed selectivity component represents the signal that is not described by a linear mixing model (Supplementary Methods M.2). The collapse in dimensionality that predicts the errors of the animal is specifically due to a weakening of the nonlinear component of mixed selectivity. Indeed, the difference in dimensionality between correct and error trials still remains significant after subtracting the linear component of mixed selectivity (Fig. 5c), whereas it is negligible when the nonlinear mixed selectivity component is removed (Fig. 5d). These results are confirmed by a principal component analysis (Supplementary Section S.17).

Discussion

We showed that the mixed selectivity that is commonly observed in PFC responses can be interpreted as a signature of high-dimensional neural representations. One advantage of high dimensionality is that information about all task-relevant aspects and their combinations is represented in a way that is easily accessible to linear classifiers, such as simple neuron models. The information is distributed across multiple neurons in an 'explicit' format[16] that allows a readout neuron to implement an arbitrary classification of its inputs. Previous studies have already shown that a linear readout is often sufficient to decode particular task aspects or to perform specific tasks (see for example refs 17, 18). Here, by showing that the neural representations are high-dimensional, we demonstrate that any binary choice task involving the 24 experimental conditions that we analysed could be performed by a linear readout.

One of our main results is that the dimensionality of the neural representations collapses in error trials, indicating that nonlinear mixed selectivity might be important for generating correct behavioural responses. It is tempting to speculate about the causes of this dimensionality collapse. Nonlinear mixed selectivity can change in a way that is compatible with our observations when neurons integrate multiple sources of information, which include those that are relevant for the task and those that are not under experimental control. The change in dimensionality may be caused by the excessive variability of sources that are not task-relevant. In other words, to perform correctly, the brain has to mix nonlinearly the task-relevant sources of information in a way that is consistent across trials. This consistency requires to restrict the contribution of the other sources. This is similar to what has been observed in the premotor cortex, where firing rates tended to be less variable on trials in which the reaction time was shorter[19]. A theoretical argument (Supplementary Section S.18) shows that neurons with a strong nonlinear mixed selectivity are more sensitive than pure selectivity neurons to the task-irrelevant sources of variability. Nonlinear mixed selectivity is most useful but also most fragile. Pure and linear mixed selectivity, which are more robust, make it possible to decode individually all task-relevant aspects even in the error trials, as observed here.

Although high dimensionality is not strictly necessary for generating rich dynamics and performing complex tasks, it is known to greatly simplify the design of local neural circuits[9]. Indeed, realizing a complex and rich dynamics is for some model circuits equivalent to solving a classification problem in which the network has to generate a particular output for each input. In these models this is typically realized by training a subset of neurons to respond in a specific way to an external input or to the internally generated activity. This is equivalent to classifying the activity of the input neurons for every time step. In many situations this activity is read out by downstream circuits. In others it is fed back to the neural circuit to affect its dynamics and hence the statistics of future inputs. Especially in the latter situations, the number of input–output functions or classifications that must be implemented by each neuron can be significantly larger than the number of functions required to simply produce the observed final behavioural response, because the neurons are required to generate the proper output for every time step. For this reason, it is often necessary to expand the dimensionality of the neuronal representations of the external sensory input and the internal state. In recent models[5–10], the dimensionality of the neuronal representations is expanded by mixing in a nonlinear way the different sources of information in a population of randomly connected neurons. The resulting neuronal representations are high-dimensional (see for example ref. 20), like those observed in PFC, and consistent with high dimensionality, the neurons show mixed selectivity which is diverse across time (that is, in different epochs of the trials) and space (that is, across different neurons). Random connectivity in small brain regions has been suggested on the basis of anatomical reconstructions[21] and recently observed in the connections from the olfactory bulb to the olfactory cortex[22] (see also ref. 14 for a general discussion).

We showed that the recorded mixed selectivity can be useful for the activity to be linearly read out. It is legitimate to ask whether these considerations would still be valid if we consider more complex nonlinear readouts. For example, some of the transformations which increase the dimensionality of the neural representations could be implemented at the level of individual neurons by exploiting dendritic nonlinearities. Our results do not exclude the functional importance of such dendritic processes. They do, however, tend to argue against a scenario where all important nonlinear transformations are carried out at the level of single neurons, a situation where dimensionality expansion could happen in a 'hidden way', and the observable representations provided by the neuronal firing rates could therefore be low-dimensional.

Finally, the particular form of redundancy inherited from high-dimensional representations allows neural circuits to flexibly and quickly adapt to execute new tasks, just as it allows them to implement

arbitrary binary classifications by modifying the weights of a readout neuron (using, for instance, a supervised procedure like the perceptron learning rule[23]). In Supplementary Section S.9 we show an example of this flexibility by training a simulated neuron to perform a new virtual task based on the recorded activity. High dimensionality might therefore be at the basis of the mechanisms underlying the remarkable adaptability of the neural coding observed in the PFC[13] and, as such, be an important element to answer fundamental questions that try to map cognitive to neurophysiological functions.

In conclusion, the measured dimensionality of the neural representations in PFC is high, and errors follow a collapse in dimensionality. This provides us with a motivation to shift the focus of attention from pure selectivity neurons, which are easily interpretable, to the widely observed but rarely analysed mixed selectivity neurons, especially in the complex task designs that are becoming progressively more accessible to investigation.

METHODS SUMMARY

The formal definitions of dimensionality and mixed selectivity are in Supplementary Methods M.1 and M.2, respectively. The procedures for removing selectivity, decoding task-relevant aspects and resampling neurons used in Fig. 3 are explained in Supplementary Methods M.3, M.5 and M.6. The dimensionality estimate of Fig. 4 is detailed in Supplementary Methods M.7. The analysis of the linear and nonlinear components of Fig. 5 is in Supplementary Methods M.8.

1. Asaad, W. F., Rainer, G. & Miller, E. K. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* **21,** 1399–1407 (1998).
2. Mansouri, F. A., Matsumoto, K. & Tanaka, K. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin card sorting test analog. *J. Neurosci.* **26,** 2745–2756 (2006).
3. Warden, M. R. & Miller, E. K. Task-dependent changes in short-term memory in the prefrontal cortex. *J. Neurosci.* **30,** 15801–15810 (2010).
4. Buonomano, D. V. & Merzenich, M. M. Temporal information transformed into a spatial code by a neural network with realistic properties. *Science* **267,** 1028–1030 (1995).
5. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14,** 2531–2560 (2002).
6. Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304,** 78–80 (2004).
7. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Rev. Neurosci.* **10,** 113–125 (2009).
8. Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63,** 544–557 (2009).
9. Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J. & Fusi, S. Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Front. Comput. Neurosci.* **4,** 24 (2010).
10. Pascanu, R. & Jaeger, H. A neurodynamical model for working memory. *Neural Netw.* **24,** 199–207 (2011).
11. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20,** 273–297 (1995).
12. Warden, M. R. & Miller, E. K. The representation of multiple objects in prefrontal neuronal delay activity. *Cereb. Cortex* **17** (Suppl. 1), i41–i50 (2007).
13. Duncan, J. An adaptive coding model of neural function in prefrontal cortex. *Nature Rev. Neurosci.* **2,** 820–829 (2001).
14. Yuste, R. Dendritic spines and distributed circuits. *Neuron* **71,** 772–781 (2011).
15. Machens, C. K., Romo, R. & Brody, C. D. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J. Neurosci.* **30,** 350–360 (2010).
16. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73,** 415–434 (2012).
17. Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K. & Poggio, T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J. Neurophysiol.* **100,** 1407–1419 (2008).
18. Klampfl, S., David, S. V., Yin, P., Shamma, S. A. & Maass, W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J. Neurophysiol.* **108,** 1366–1380 (2012).
19. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* **26,** 3697–3712 (2006).
20. Barak, O., Rigotti, M. & Fusi, S. The sparseness of mixed selectivity neurons controls the generalization–discrimination trade-off. *J. Neurosci.* **33,** 3844–3856 (2013).
21. Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* 2nd edn (Springer, 1998).
22. Sosulski, D. L., Bloom, M. L., Cutforth, T., Axel, R. & Datta, S. R. Distinct representations of olfactory information in different cortical centres. *Nature* **472,** 213–216 (2011).
23. Rosenblatt, F. *Principles of Neurodynamics* (Spartan Books, 1962).

Author Contributions M.R.W. and E.K.M. performed the experiments and collected the data. M.R., O.B., X.-J.W. and S.F. developed the theoretical framework. M.R., O.B., N.D.D. and S.F. conceived the data analyses. M.R. performed the data analyses. M.R., O.B. and S.F. wrote the paper.